

Documentação Sprint 3

Grupo MAI

DISRUPTIVE ARCHITECTURES: IOT, IOB & GENERATIVE IA

Equipe

Luiza Nunes de Jesus - RM99768

Melissa de Oliveira Pecoraro - RM98698

Pedro Marques Pais Pavão - RM550252

Pamella Schimalesky Engholm - RM551600

Roberto Menezes dos Santos - RM552520

Proposta

A MAI, empresa criada para a resolução do problema proposto, tem como objetivo criar uma inteligência artificial que estude o comportamento do cliente através de suas compras, listas de desejo, pesquisas, entre outros, e possibilita que as empresas gerem uma publicidade específica para o perfil de cada cliente.

Para que essa publicidade seja eficiente, a MAI acrescentará ao aplicativo da empresa uma timeline de vídeos, estes que serão disponibilizados conforme o perfil do consumidor. Os vídeos serão previamente criados para o marketing de marcas presentes no aplicativo.

Além disso, postagens em formato de carrossel serão disponibilizadas para os consumidores que desejam informações rápidas no próprio aplicativo da empresa, como tendências de roupas, cores da estação, etc.

Desenvolvimento da solução

No começo apresentamos a nossa ideia, qual o foco do nosso projeto, o que faríamos e colocamos isso em prática desenvolvendo um modelo de machine learning.

Neste momento trazemos a demonstração do protótipo funcional e as funcionalidades implementadas até agora, detalhando a arquitetura de IA.

Como já foi falado, nosso objetivo é o varejo online e filtramos nessa etapa o foco na segmentação de clientes, entendendo seus comportamentos e analisando insights.

Para isso, utilizamos uma carga de dados sobre compras de clientes no varejo online e fizemos o pré-processamento de dados, uma análise exploratória e utilizamos o K-means, um algoritmo de clusterização, para segmentação dos clientes, analisando grupos com características semelhantes. Além disso foi desenvolvido um dashboard para visualização do que foi analisado, um modelo de classificação para classificar os grupos de clientes possíveis e para melhorar o desenvolvimento da solução.

Também foi feita uma integração com o Flask, tornando o protótipo mais funcional e adequado a realidade do projeto, sendo possível consumir o que foi desenvolvido, a api, em qualquer lugar ou aplicação.

A arquitetura escolhida de Machine Learning é a escolha mais robusta e flexível para a segmentação de clientes no varejo online. Ela permite uma segmentação precisa, automatizada e escalável, oferecendo insights valiosos que podem ser utilizados para estratégias de marketing mais eficazes e personalizadas. Em comparação com outras abordagens, essa arquitetura é superior em termos de adaptabilidade, eficiência, capacidade de lidar com grandes volumes de dados e a possibilidade de integração com outros sistemas, apis etc., tornando-a a solução ideal para atender às necessidades do seu projeto.

O fato também de avançarmos para o algoritmo de classificação depois de fazer a segmentação, nos mostra que conseguimos uma melhor predição e identificação de clientes e seus respectivos grupos, sendo a melhor solução para a ideia.

Frameworks/Bibliotecas

Para chegarmos nos resultados e entendermos sobre a segmentação, utilizamos a biblioteca pandas que nos ajuda a carregar o dataset utilizado, a limpar dados para deixarmos nosso dataset padronizado e transformar os dados, como o que foi feito com a coluna 'InvoiceDate' e manipulação dos dados/colunas nos permitindo analisar mais profundamente as médias de compras, frequências, valor total gasto por clientes, etc.

Para a análise exploratória utilizamos o matplotlib ou seaborn, com eles conseguimos visualizar os dados e aquilo que estamos manipulando, através de gráficos e insights sobre características dos clientes, sendo possível chegarmos a muitas conclusões através do que eles proporcionam, como a distribuição das compras por ano, produtos mais vendidos de acordo com determinado ano, o entendimento da relação entre quantidade e preço unitário, os principais países que realizam pedidos, os clientes que tem o maior gasto total, etc.

Utilizamos também o scikitlearn, sendo utilizado no nosso código: 'from sklearn.cluster import KMeans', para mexermos com o algoritmo de K-means, para treinarmos modelos e analisarmos a segmentação de clientes.

Para fazermos o modelo de classificação importamos por exemplo o RadomForest, o algoritmo utilizado para essa solução, e também para a integração com o Flask, importamos o pickle.

Conceitos de Machine Learning

Nessa segunda etapa, utilizamos conceitos de machine learning que identificam grupos de clientes com características semelhantes, segmentando clientes.

Para isso, o algoritmo indicado para ter essa resposta é o K-means, que aprende a organizar itens parecidos juntos. Ele faz isso transformando cada item do conjunto de dados em um ponto em um espaço de muitas dimensões. O número de dimensões é igual ao número de características de cada item.

Esse algoritmo está relacionado ao conceito de clusterização/agrupamento, do aprendizado não supervisionado, que analisa os dados e visa agrupar objetos similares em grupos ou clusters.

Com o desenvolvimento conseguimos chegar a 3 segmentos distintos da nossa base de dados:

- os compradores frequentes econômicos;
- os compradores moderados;
- os compradores esporádicos de luxo.

Baseado na segmentação que conseguimos desenvolver, levamos isso para um modelo de classificação, sendo possível trazer mais praticidade e melhor desenvolvimento para aquilo que temos como objetivo, então é utilizada a clusterização, mas depois desse processo ser muito bem definido, utilizamos a classificação, que possui como resposta um desses 3 tipos de grupos de clientes.